Anomaly Detection in Crowded Scenes Based on Group Motion Features

Shuqiang Guo, Dongxue Li, Lili Yao

School of Computer Science, Northeast Electric Power University, China
guoshuqiang@gmail.com, baiqianlong@163.com, yaolili@163.com

Abstract

Event detection in crowded scenes is a challenging task for Computer Vision. In this study, based on group motion features, we propose an approach for crowded scene anomaly detection and localization. According to the motion trajectory of numerous pedestrians, both distance and relative speed between trajectories can be extracted, and the pedestrian groups can be recognized via their spatial relationship. Anomaly events in crowded scenes can be detected based on variations of group numbers and speed. To demonstrate the effectiveness of the approach, a quantitative experimental evaluation has been conducted on multiple, publicly available video sequences.

Keywords: Anomaly detection, Crowded scenes, Group motion features, SVM

1 Introduction

Video based anomaly detection can provide a powerful guarantee for the early warning of public security and crowd evacuation. If the existing monitoring system can become sufficiently intelligent to analyze the behavioral state of group events and judge whether abnormal events exist (such as crowd trampling, fights, or disturbances), disaster events can be avoided or at least effectively detected to set off an alarm. At present, the state detection of crowds still requires significant human labor in the field of public security. Automatic detection methods based on machine vision and artificial intelligence are also in constant development. During recent years, the detection method of the abnormal state of crowds has been mainly divided into the following three categories:

1.1 Anomaly Detection Based on Pixel Statistics

The earliest accessible feature is the pixel statistical feature, and it is still widely used for measuring crowd density. The basic idea about this algorithm is using the number of pixels or edge points of the foreground pixels of a crowd to represent the crowd density. Moreover, whether these pixel features can be used to estimate the crowd density depends on the validity of the background model or the edge detection algorithm. Davies is a pioneer in the use of computer vision to study the crowd information, and he applied such a crowd density monitoring system to the London Metro station [1]. Davies et al. assumed the foreground pixels, or edge pixel number, to have a linear relationship with the number of crowds. The crowd density is then evaluated using pixel statistics after both background subtraction algorithm and edge detection; however, the accuracy of this algorithm decreases when the crowds overlapped. In 2005, Ma et al. applied the perspective effect correction algorithm to the crowd monitoring system [2]. In this method, foreground pixels of the image are differently weighted, but it still lacks high accuracy. In 2006, Kong et al. standardized crowd characteristics according to the point of view of the camera, and the training set did not require to be changed when the camera was reinstalled [3]. Xiaohua and Pathan used Support Vector Machine (SVM) to first analyze the crowd based on perspective correction [4-6]. However, these methods need to prepare a large number of positive and negative samples in advance; therefore, it is difficult to meet the requirements of real-time anomaly detection.

1.2 Anomaly Detection Based on Texture Analysis

In the 1970s, Sklansky proposed a texture feature extraction method based primarily on texture analysis [7]. He provided support for the theory and the technological accumulation for the subsequent texture research. Sklansky applied the Gray Level Dependence Matrix (GLDM) to crowd event detection. The GLDM method is a common method to describe the texture by studying spatial correlation properties of the gray scale, in which the correlation between the gray levels is reflected by two pixels of an image within a certain distance. GLDM is a common method to describe the texture by studying the spatial correlation properties of the gray scale. Commonly used statistics based on the

*Corresponding Author: Shuqiang Guo; E-mail: guoshuqiang@gmail.com
DOI: 10.3966/160792642020052103024

GLDM are Contrast, Homogeneity, Energy, and Entropy [8-13]. Because it can reduce impact of occlusion, these types of statistical features can be used to detect a crowd anomaly event. However, the detection validity of this method is inferior for low crowd density.

1.3 Anomaly Detection Based on Optical Flow

In 2009, Mehran et al. proposed an anomaly detection model for moving crowds using a social force model [14]. This model extracts dynamic information, including the velocity and direction of crowds, and represents the individual and mutual movement of crowds via optical flow. During recent years, further optical flow methods, combined with modern theory, have been used for the detection of abnormal people, such as neural networks [15-19], deep learning [20-21], and the topological method [22-24]. These methods can effectively show the details of crowd dynamics; however, due to its computational complexity, they are very time consuming.

Currently, new methods for abnormal detection in crowd scenes are continuously described. Greco et al. proposed a solution, which is able to recognize simple and complex events, starting from the output of a people-tracking algorithm [25-26]. Bera et al. proposed an algorithm for real-time anomaly detection in low to medium density crowd videos using trajectory-level behavior learning [27]. Brun et al. defined a novel and efficient kernel-based clustering algorithm, aimed at obtaining groups of normal trajectories [28]. This study proposes a crowd anomaly detection algorithm based on crown density and velocity features. In this method, the background subtraction method has been used to track pedestrians. According to trajectory, pedestrians were divided into many groups, and then the abnormal crowd state could be detected using the speed of each group and alterations of group number. Experimental results show that the method proposed in this paper can effectively and with high accuracy detect the crowd anomaly state.

2 Methodology

2.1 Overview

The utilized processing flow of crowd anomaly detection is shown in Figure 1. Firstly, the motion trajectory of pedestrians could be obtained based on the mean-shift method. According to the motion trajectory, pedestrians were divided into numerous groups. Finally, the abnormal state of pedestrians could be detected based on group features.

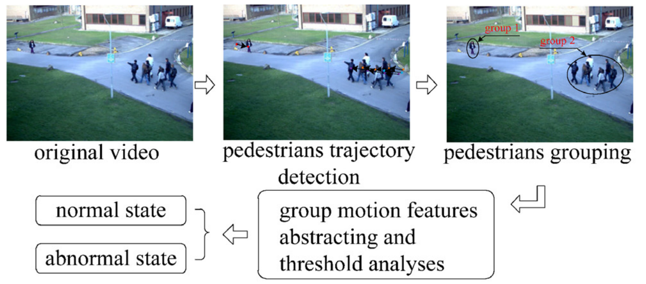


Figure 1. Processing procedure of crowd anomaly detection

2.2 Motion Trajectory Detection of Pedestrians Based on Mean-Shift

First proposed by Fukunaga and Hosteter in 1975, the mean-shift algorithm is a method for data analysis based on kernel density estimation [29-30]. After 1995, it was widely used in image segmentation, image smoothing, medical image analysis, object tracking, and other image processing fields [31]. The mean-shift algorithm uses a color histogram, sets the weight coefficient according to the contribution to the mean shift vector of a sample point, and then obtains the maximum offset of the solution via the mountain climbing algorithm to track a non-rigid target.

According to the Mean-shift algorithm, the weight of a pixel at position x can be obtained via Equation 1.

$$w(x)=\sum_{u=1}^{m}\sqrt{\frac{q_u}{p_u(x_0)}}\delta\left[b(x)-u\right] \tag{1}$$

where q_u is the normalized weighted color histogram of target mode, $p_u(x_0)$ is the normalized weighted color histogram of searching region, whose center is x_0, $b(x)$ is the bin number of gray value in position x, u is the bin number, and δ is Kronecker delta function.

The moving vector is then given as:

$$\hat{p}_u(x)=\frac{\sum_{u=1}^{m}(1-hx_i)\delta\left[b(x_i^*)-u\right]}{\sum_{u=1}^{m}(1-hx_i)} \tag{2}$$

where $K(x)$ represents the Kernel function of the normal distribution. The iteration stops when the moving vector is below threshold.

When the Mean-shift algorithm is applied to object tracking, it is necessary to determine the object mode in the initial position. For our method, the target human windows in the initial position were determined via differences from the background image, as shown in Figure 2.

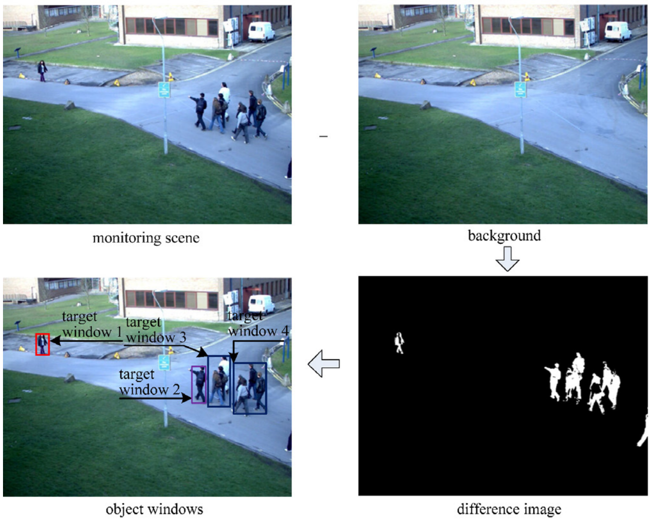


Figure 2. Target window searching

The normalized color histogram q_u could be calculated according to the target windows. The algorithm for human motion trajectory is listed as follows:

(1) According to the background difference method to determine the number of tracking target and the size of the target window.

(2) To remove the influence of the target scale, it is normalized to a unit circle and the target model { }(u =1 2…m, m bins of histograms) is derived from a target window centered at x0 with its pixel coordinates$\{x_i^*\}$.

$$\hat{q}_u = \frac{\sum_{u=1}^{m}(1-\|x_i^*\|^2)\delta\left[b(x_i^*)-u\right]}{\sum_{u=1}^{m}(1-\|x_i^*\|^2)} \quad (3)$$

where n represents the number of pixels, δ represents the Kronecker delta function:

$$\delta(x) = \begin{cases} 1 & x = 0 \\ 0 & otherwise \end{cases} \quad (4)$$

To estimate { $\hat{p}_u(\hat{x}_0)$ } in the new frame, x0 from the previous frame location is used as the initial position.

$$\Delta x = \frac{\sum_{i=1}^{m} K(x_i - x_0)\omega(x_i)(x_i - x_0)}{\sum_{i=1}^{m}\left|K(x_i - x_0)\omega(x_i)\right|} \quad (5)$$

(3) If $\|\hat{x}_1 - \hat{x}_0\| < \varepsilon$ stop. Otherwise, use the current target location as a start for the new location $\hat{x}_1 = \hat{x}_0$, and continue with Step 3.

(4) If the target model { $\hat{q}_u$ }is the last objects, stop. Otherwise, go to Step 1.

(5) Derive weights for i = 1, 2, …, n according to Equation 1.

(6) Determine the new location of the target candidate according to Equation 2.

2.3 Group Detection Based on Human Motion Trajectories

According to the aforementioned human motion trajectories, it is possible to determine whether the pair trajectories belong to one group, and whether this group is comprised of three or more persons.

2.3.1 Spatial-temporal Feature of the Pair Trajectories

To judge whether the pair trajectories are in one group, the distance between two targets, the direction of velocity of both targets, and other features are proposed, as shown in Figure 3. The four Spatial-temporal features are expressed via F_1, F_2, F_3, and F_4. Their definitions are given as:

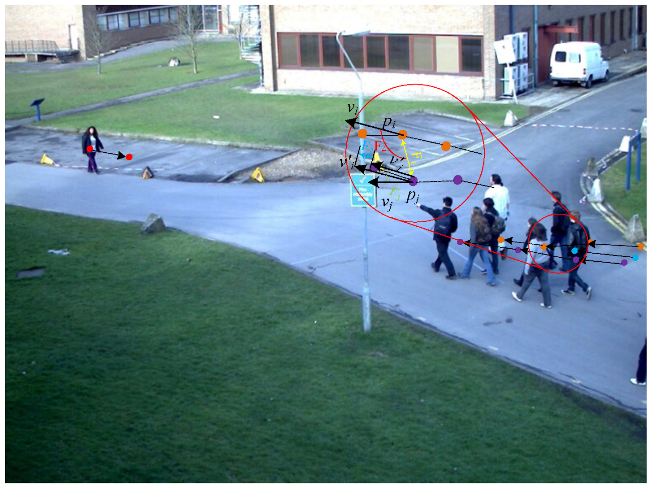


Figure 3. Spatial-temporal feature abstracting

$F_1 = |p_i - p_j|$, where p represents the position of target i, F1 represents the distance of two targets.

$F_2 = |v_i| - |v_j|$, where v_i represents the velocity of target i, F_2 represents the difference between the velocity of both targets.

$F_3 = |\arctan(v_i) - \arctan(v_j)|$, where $\arctan(v_i)$ represents the direction between two targets, F_3 represents the difference in the direction of both targets.

$F_4 = |\arctan(p_i - p_j) - \arctan(v_i)|$, F_4 is the difference between the direction of velocity of one target and the direction of relative position of both targets.

These four spatial-temporal features are calculated for every frame, and they can be expressed via histograms every 30 frames, respectively, as shown in Figure 4. Then, the feature vectors of every trajectory pair could be obtained. According to these feature vectors, whether both pedestrians are in one group can be judged using SVM.

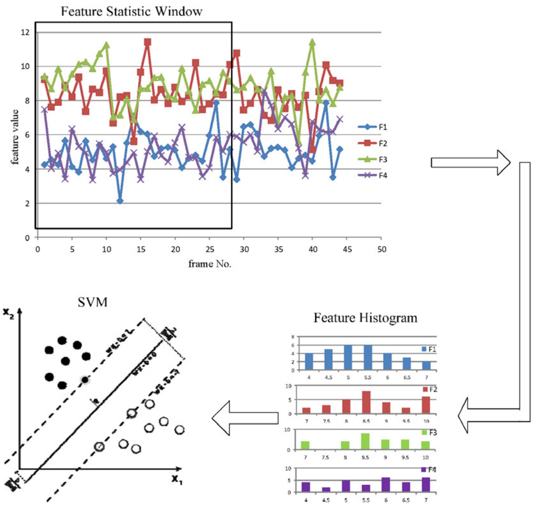


Figure 4. Feature statistic method

2.3.2 Target Pair Detection in one Group Based on Spatial-temporal Features

Using the spatio-temporal vector defined in the previous section, pedestrian pair detection in one group is conducted in all trajectory pairs. As shown in Figure 5, a pedestrian pair in one group is first detected within every frame. If the detection rate of all frames is above the threshold value, the pedestrian pair is considered to be in one group. At the time of machine learning, the spatio-temporal features within group and other groups are considered as positive and negative examples in SVM.

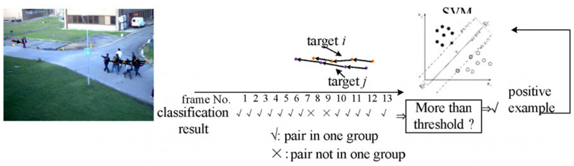


Figure 5. One group judgment with pair targets

Since pair detection in one group is based on spatial-temporal features of all frames within the video, this does not impact the overall recognition rate even if part of the frames cannot be effectively recognized. However, if the deviance of learning samples is small, i.e. the spatial-temporal vector is near the identification boundary, pedestrian pair detection is difficult. To solve this problem, we improved the method to detect the pedestrian pair using Bag of Features. In this method, the pedestrian pair detection in one group is based on the distribution (histogram) of the spatio-temporal feature in the trajectory of 20 frames. At the time of machine learning, spatio-temporal vectors of all frames are first clustered and the histogram can then be obtained by designating the feature vector in the same cluster to the same bin. A histogram regarded as the feature vector of the trajectory pair is used as a sample set to train the SVM. In this method of judgment, similar vectors are appointed to the same bin; therefore, subtle variations in the frames do not have a significant effect on the decision regardless of whether the trajectory pair belongs to one group. Furthermore, since the SVM is applied to the histogram of the feature vector of the frames, proper clustering requires to identify the data points that are matched to the occurrence frequency of similar feature vectors in all frames.

2.3.3 Group Detection with More than Three Targets

As mentioned in Section 2.2, the group with pair targets was detected. As shown in Figure 6, if the same target is found in different groups, the groups were merged into one group. Therefore, a group with more than three targets could be detected.

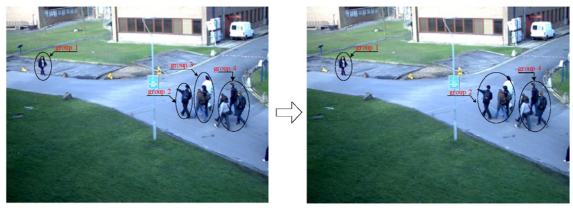


Figure 6. One group judgment more than three targets

2.4 Anomalous Event Judgment

When the event occurs, the movement type of the crowd mainly includes three categories: rapid division of the crowd, rapid merger of the crowd, and rapid acceleration of the crowd. According to these three types of movement, the feature vectors can be summarized as follows:

$$X = \begin{bmatrix} \left| \dfrac{\partial^2 N_g}{\partial t^2} \right| \\ \dfrac{\partial^2 P_g}{\partial t^2} \end{bmatrix} \tag{6}$$

Where Ng represents the number of groups detected by the method mentioned in Section 2, and $\left| \dfrac{\partial^2 N_g}{\partial t^2} \right|$ represents both the rapid division and the rapid merger of the crowd. P_g represents the position of the group and $\dfrac{\partial^2 P_g}{\partial t^2}$ represents the rapid acceleration of the crowd. Using the feature vector X obtained from each frame, an abnormal event can immediately be detected if the difference of the feature vector between two frames is larger than the threshold value.

3 Experiments

This section provides a detailed introduction of the entire experiment, including the testing environment, the evaluation method, and the experiment itself. The end of this section provides an analysis of algorithm performance.

3.1 Group Detection Based on Human Motion Trajectories

Two different data sets have been selected for experiments: UMN and PETS2009. Each data set contains one or multiple video sequences and the corresponding ground truth data. The UMN data set has been collected by the University of Minnesota, USA, and consists of eleven videos that represent escape events. The videos were captured in three different indoor and outdoor settings, commonly denoted as Lawn, Indoor, and Plaza. Each video starts with a crowd, of about 20 people, walking in different directions; then, an abnormal event causes people to run away. PETS 2009 has been recorded for the workshop PETS 2009 at the Whiteknights Campus of the University of Reading, UK. PETS 2009 comprises multi-sensor sequences containing crowd scene scenarios with increasing scene complexity and consists of three data sets: (S1) concerning person count and density estimation; (S2) addressing people tracking; (S3) involving flow analysis and event recognition. In our experiments, we used the S3 data set (see Figure 8). The experiments were implemented on a computer with an Intel i3 processor 2.7 GHz CPU, with a 4G memory capacity.

3.2 Evaluation Method

The Recall Rate, Precision Rate, Real-time Performance, True Positive Rate (TPR), and False Positive Rate (FPR) were utilized to evaluate the validity of our method. In the video sequences, TP was defined as the number of abnormal events correctly detected, FP represents the number of normal events that are detected as abnormal events, TN is defined as the number of normal events that were correctly detected, FN represents the number of abnormal events that were detected as normal events and the definitions of the evaluation items are given as follows:

(1) Recall Rate: The recall rate is equal to the number of successfully detected abnormal events divided by the total number of abnormal events.

$$Recall\ Rate = \frac{TP}{TP+FN} \qquad \textbf{(7)}$$

(2) Precision Rate: The precision rate expresses the accuracy of our detection algorithm. It is equal to the number of correctly detected abnormal events divided by the total number of detected abnormal events.

$$Precision\ Rate = \frac{TP}{TP+FP} \qquad \textbf{(8)}$$

(3) Real-time Performance: The real time performance is one of the most important evaluation indicators of an intelligent video surveillance system. This is attributable to the fact that the monitor has to be noticed as soon as the abnormal event occurs. In this study, the time complexity of our algorithm was defined as the ratio of the processing time of all frames and the actual playing time of the video.

(4) True Positive Rate: This is identical to the Recall Rate, and it could be computed with Equation 7.

(5) False Positive Rate: The false positive rate expresses the rate of the number of normal events that are registered as abnormal events to all abnormal events. It can be computed with the following formula:

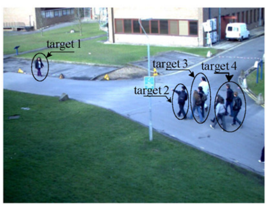
$$FPR = \frac{FP}{FP+TN} \qquad \textbf{(9)}$$

TPR and FPR can be used to generate a Receiver Operating Characteristics (ROC) curve to evaluate the performance of the anomaly detection algorithms.

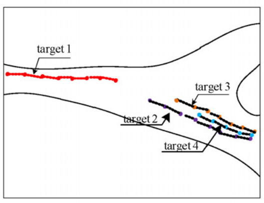
3.3 Experiment and Result Analysis

3.3.1 Group Detection Experiment

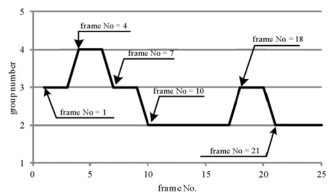
The experimental data of this study partially originated from PETS 2009. We received the original video as shown in Figure 7(a). The pedestrians in the video were grouped based on the method described in Section 2.2. The results obtained are shown in Figure 7(b) and Figure 7(c). Figure 7(b) displays the detection results for the target trajectory combining mean-shift and background frame difference algorithms. The grouping results are displayed in Figure 7(c) and are based on target trajectories. According to Figure 7(c), there were three targets (target 1, target 2, and target 3) in the first frame, and they were initially divided into three groups. In the 4th frame, target 4 entered the camera view range. Because the spatio-temporal features of target 2 and target 3 were similar, they were merged into one group in the 7th frame, changing the number of groups from 4 to 3. In the 10th frame, target 2, target 3, and target 4 merged into one group, changing the group number to 2.



(a) A frame in video



(b) Detection result of targets trajectory



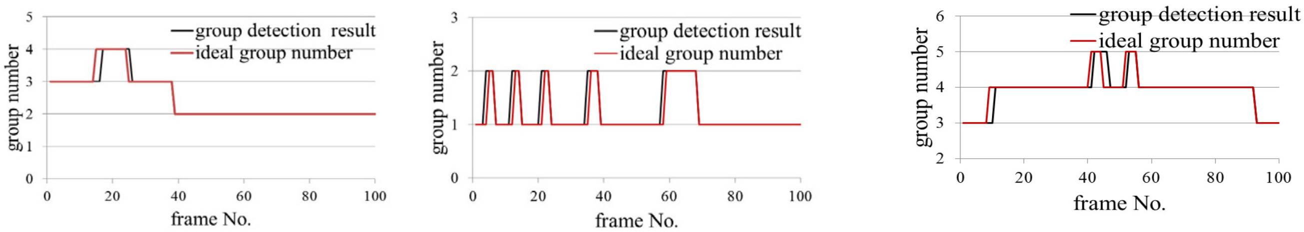
(c) Grouping result of targets

Figure 7.

To qualitatively evaluate the performance of our grouping method, we tested the approach using the videos of sparse crowd, medium-density crowd, and high-density crowd from the PETS2009 data set. One frame of the original videos is shown in Figure 8. The group detection experiments were conducted using the three videos respectively, and the results are shown in Figure 9. As shown in Figure 9, the black solid line expresses the group number detected by our algorithm, while the red solid line expresses the ideal group number. It can be seen that the group number detected by our algorithm is almost identical to the ideal group number. The accuracy of sparse crowd, medium-density crowd, and high-density crowd were 96%, 95%, and 94%. Moreover, the accuracy was closely related to the crowd density and the times of grouping transformation. The larger the crowd density and the times of grouping transformations were, the smaller the accuracy would become.

(a) Sparse crowd (b) Medium density crowd (c) Dense crowd

Figure 8. Original dataset



(a) Sparse crowd (b) Medium density crowd (c) Dense crowd

Figure 9. Group detection result

3.3.2 Anomaly Detection Experiment

To achieve representativity of the experimental results of the anomaly detection algorithm proposed in this paper, we used standard UMN and PETS2009 data set to test our algorithm. Figure 10 exhibits one sample frame of each of the considered sequences. For performance comparison, we chose four state-of-the-art methods: the Energy Model (EM) [8], Cell- based Analysis (CA) [12], Graph Modeling and Matching (GMM) [22], and Density Estimation (DS) [23]. The quantitative results of these for methods were partially obtained from the source code management functionality Github, and partially from the proceeding on computer vision. The abbreviation of our proposed method is Group Motion Features (GMF).

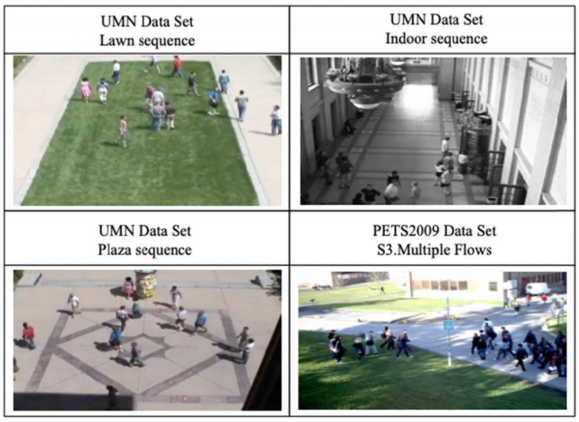


Fig. 10. Normal frames in UMN Data Set

With regard to the frame-level measurements, Figure 11 shows the ROC curves of the competing models. For the frame-level ROC, the proposed algorithm performs better than others except for the crowd GMM where the false positive rate was high. Based on the ROC curves, Recall Rate, Precision Rate, and Real-time values are listed in Table 1.

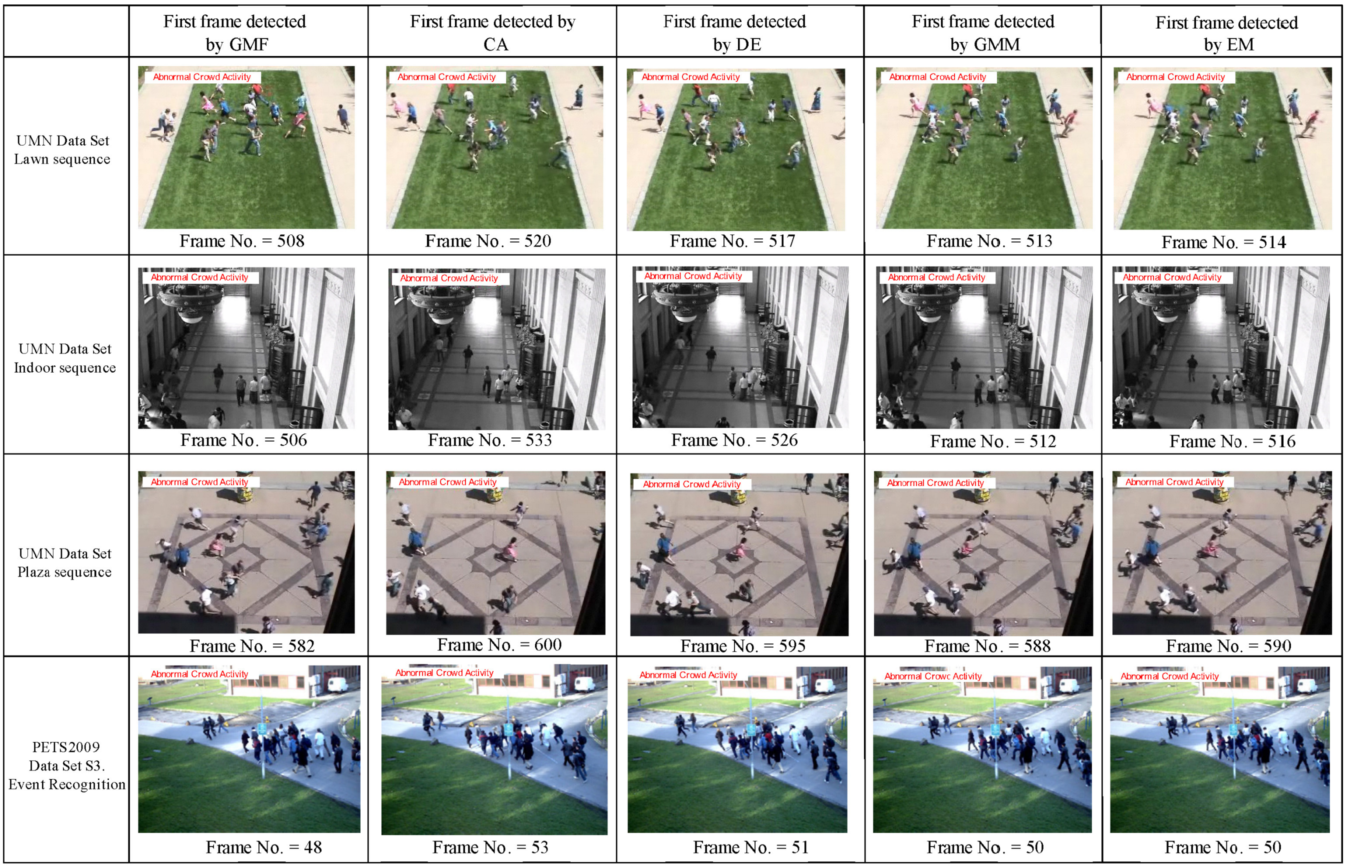


Fig. 11. The first frame of abnormal event detection using different method

Table 1. Performance comparison in UMN data set

sequence	method	Recall ratio	Precision ratio	Real-time
Lawn sequence	DS	92.8%	85.2%	1.9
	CA	87.3%	83.8%	4.3
	EM	90.2%	84.1%	3.2
	GMM	95.6%	87.2%	2.8
	GMF	97.4%	89.3%	5.5
Indoor sequence	DS	90.1%	82.4%	1.5
	CA	88.5%	83.1%	4.8
	EM	89.3%	80.9%	3.9
	GMM	94.8%	85.5%	3.2
	GMF	95.6%	87.7%	4.9
Plaza sequence	DS	94.2%	81.4%	2.1
	CA	91.2%	78.8%	5.0
	EM	92.5%	81.8%	4.3
	GMM	94.6%	83.5%	3.4
	GMF	96.8%	85.3%	5.3
S3 Multiple Flows	DS	90.6%	83.6%	1.8
	CA	89.9%	80.3%	4.6
	EM	92.6%	83.6%	3.9
	GMM	93.4%	82.4%	3.3
	GMF	94.3%	84.4%	4.9

Judging from these numbers, the recall and precision feature of the detection of events and the aggregation of the four types of group events was high, the complex background of the scene was also applicable. In addition, the algorithm accurately detected moving targets in the foreground, as well as eliminating interference caused by noise and shadow, thus greatly improving detection efficiency.

However, the required calculation was significant because the algorithm needs to spend a lot of time on feature extraction; consequently, it is therefore not able to beat CA, DE, GMM and EM in real time. If frame detection is very difficult to achieve to meet real-time requirements, there is need to improve the algorithm. For example, to meet the requirement of real-time detection, every 10 frames would be in need of feature processing, or the feature extraction algorithm has to be optimized.

Moreover, as show in Figure 12, an abnormal event was detected at frame number 508 using our method; and when using CA, DE, GMM and EM method in the lawn sequence, the first abnormal frame numbers were 520, 517, 513 and 514. For the other two sequences, the frame number of the first abnormal event detected by our method was smaller than for other methods. This supports that our proposed method is more accurate in determining the time of the anomaly event.

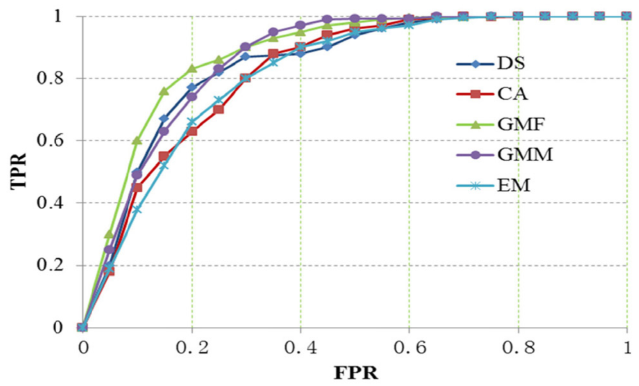


Figure 12. ROC curves for PETS 2009 S3

4 Conclusions

In this study, we developed a group feature method for anomaly detection and localization in different scenes obtained via UMN database sets. Our method first extracts the trajectories of pedestrians, and pedestrians are then grouped according to the trajectories in the image. By using the group feature of pedestrians, their abnormal behavior can be detected. The detection result of our method is superior to that of methods based on the population density because it considers the mutual relationship between pedestrians. Experiments on four different types of benchmarks show that the performance of our method is better than that of state-of-the-art approaches, particularly on anomaly event localization. However, the algorithm needs to utilize a significant amount of time for feature extraction. Therefore, it is necessary to optimize the algorithm to improve the speed of feature extraction. In addition, the threshold of anomaly detection was difficult to determine; therefore, in future work, a method of adaptive threshold should be developed to detect abnormal events.

Acknowledgments

This research is partially supported by:
(1) Research Foundation of Education Bureau of Jilin Province(JJKN20190710KJ).
(2) Research Foundation of Education Bureau of Jilin Province (JJKH20180447KJ)

References

[1] A. C. Davies, J. H. Yin, S. A. Velastin, Crowd Monitoring Using Image Processing, *Electronics Communication Engineering Journal*, Vol. 7, No. 1, pp. 37-47, February, 1995.

[2] R. Ma, L. Li, W. Huang, Q. Tian, On Pixel Count Based Crowd Density Estimation for Visual Surveillance, *IEEE Conference on Cybernetics and Intelligent Systems*, Singapore, Singapore, 2004, pp. 170-173.

[3] D. Kong, D. Gray, H. Tao, A Viewpoint Invariant Approach for Crowd Counting, *18th International Conference on Pattern Recognition (ICPR'06)*, Hong Kong, China, 2006, pp. 1187-1190.

[4] X. Li, L. Shen, H. Li, Estimation of Crowd Density Based on Wavelet and Support Vector Machine, *Transactions of the Institute of Measurement and Control*, Vol. 28, No. 3, pp. 299-308, August, 2006.

[5] S. Pathan, A. Ayoub, M. Bernd, Incorporating Social Entropy for Crowd Behavior Detection Using SVM, *International Symposium on Visual Computing*, Las Vegas, NV, USA, 2010, pp. 153-162.

[6] F. Sun, Z. Zhang, Y. Kao, T. Li, B. Shen, A New Method to Detect the Adversarial Attack Based on the Residual Image, *Journal of Internet Technology*, Vol. 20, No. 4, pp. 1297-1304, July, 2019.

[7] J. Sklansky, Image Segmentation and Feature Extraction, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 8, No. 4, pp. 237-247, April, 1978.

[8] A. N. Marana, S. A. Velastin, L. F. Costa, R. A. Lotufo, Automatic Estimation of Crowd Density Using Texture, *Safety Science*, Vol. 28, No. 3, pp. 165-175, April, 1998.

[9] G. Xiong, J. Cheng, X. Wu, Y. Chen, Y. Ou, Y. Xu, An Energy Model Approach to People Counting for Abnormal Crowd Behavior Detection, *Neurocomputing*, Vol. 83, pp. 121-135, April, 2012.

[10] M. J. V. Leach, E. P. Sparks, N. M. Robertson, Contextual Anomaly Detection in Crowded Surveillance Scenes, *Pattern Recognition Letters*, Vol. 44, pp. 71-79, July, 2014.

[11] A. Fagette, N. Courty, D. Racoceanu, J. Dufour, Unsupervised Dense Crowd Detection by Multiscale Texture Analysis, *Pattern Recognition Letters*, Vol. 44, pp. 126-133, July, 2014.

[12] V. Reddy, C. Sanderson, C. B. Lovell, Improved Anomaly Detection in Crowded Scenes via Cell-based Analysis of Foreground Speed, Size and Texture, *CVPR 2011 WORKSHOPS*, Colorado Springs, CO, USA, 2011, pp. 55-61.

[13] R. Mehran, A. Oyama, M. Shah, Abnormal Crowd Behavior Detection Using Social Force Model, *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, 2009, pp. 935-942.

[14] M. Fu, P. Xu, X. Li, Q. Liu, M. Ye, C. Zhu, Fast Crowd Density Estimation with Convolutional Neural Networks, *Engineering Applications of Artificial Intelligence*, Vol. 43, pp. 81-88, 2015.

[15] A. Pennisi, D. D. Bloisi, L. Iocchi, Online Real-time Crowd Behavior Detection in Video Sequences, *Computer Vision and Image Understanding Individual and Group Activities in Video Event Analysis*, Vol. 144, pp. 166-176, August, 2016.

[16] J. Xu, S. Denman, V. Reddy, C. Fookes, S. Sridharan, Real-time Video Event Detection in Crowded Scenes Using MPEG Derived Features: A Multiple Instance Learning Approach, *Pattern Recognition Letters Pattern Recognition and Crowd Analysis*, Vol. 44, pp. 113-125, July, 2014.

[17] R. Chaker, Z. A. Aghbari, I. N. Junejo, Social Network Model for Crowd Anomaly Detection and Localization,

Pattern Recognition, Vol. 61, pp. 266-281, January, 2017.
[18] S. Zhou, W. Shen, D. Zeng, M. Fang, Y. Wei, Z. Zhang, Spatial–temporal Convolutional Neural Networks for Anomaly Detection and Localization in Crowded Scenes, *Signal Processing*, Vol. 47, pp. 358-368, September, 2016.
[19] Y. Feng, Y. Yuan, X. Lu, Learning Deep Event Models for Crowd Anomaly Detection, *Neurocomputing*, Vol. 219, pp. 548-556, January, 2017
[20] S. Yi, H. Li, X. Wang, Pedestrian Behavior Understanding and Prediction with Deep Neural Networks, *Computer Vision – ECCV 2016*, pp. 263- 279, September, 2016.
[21] N. Li, Z. Zhang, Abnormal Crowd Behavior Detection Using Topological Methods, *12th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, Sydney, NSW, Australia, 2011, pp. 13-18.
[22] D. Y. Chen, P. C. Huang, Visual-Based Human Crowds Behavior Analysis Based on Graph Modeling and Matching, *IEEE Sensors Journal*, Vol. 13, No. 6, pp. 2129-2138, February, 2013.
[23] I. R. Almeida, C. R. Jung, Change Detection in Human Crowds, *2013 XXVI Conference on Graphics, Patterns and Images*, Arequipa, Peru, 2013, pp. 63-69.
[24] L. Greco, P. Ritrovato, A. Saggese, Abnormal Event Recognition: A Hybrid Approach Using Semantic Web Technologies, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Las Vegas, NV, USA, 2016, pp. 1297-1304.
[25] R. K. Wang, Z. M. Lu, Video Moving Object Detection Based on Block Truncation Coding, *Journal of Information Hiding and Multimedia Signal Processing*, Vol. 8, No. 2, pp. 510-516, January, 2017.
[26] A. Bera, S. Kim, D. Manocha, Realtime Anomaly Detection Using Trajectory-Level Crowd Behavior Learning, *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Las Vegas, NV, USA, 2016, pp. 1289-1296.
[27] L. Brun, A. Saggese, M. Vento, Dynamic Scene Understanding for Behavior Analysis Based on String Kernels, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 24, No. 10, pp. 1669-1681, January, 2014.
[28] K. Fukunaga, L. Hostetler, The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition, *IEEE Transactions on Information Theory*, Vol. 21, No. 1, pp. 32-40, January, 1975.
[29] D. Comaniciu, P. Meer, Mean Shift Analysis and Applications, *Proceedings of the Seventh IEEE International Conference on Computer Vision*, Kerkyra, Greece, 1999, pp. 1197-1203.
[30] R. Wang, Z. Lu, Video Moving Object Detection Based on Block Truncation Coding, *Journal of Information Hiding and Multimedia Signal Processing*, Vol. 8, No. 2, pp. 510-516, March, 2017.
[31] D. Comaniciu, P. Meer, Mean Shift and Optimal Prediction for Efficient Object Tracking, *Proceedings 2000 International Conference on Image Processing*, Vancouver, BC, Canada, 2000, pp. 70-73.

Biographies

Shuqiang Guo rceived his M.S. degree in Computer science and technology from School of Information Engineering of Northeast Electric Power University in 2006. He received the D.S. degree in information science from Yamagata University in 2010.Since 2012, he has become a teacher of Northeast Electric Power University, and his main research interests include image processing, pattern recognition, and computer vision.

Dongxue Li is currently the Mater degree in Computer Science with Northeast Electric Power University. Her research interests are the areas of computer vision and artificial intelli- gence.

Lili Yao is currently the master degree in Computer Science with Northeast Electric Power University. Her research interests are image process- ing and computer technology.